Demography India
Vol. 50, No. 2 (2021), pp. 16-32 ISSN 0970-454X

Population Projection of Metropolitan Cities in India: An Approach to Bayesian Hierarchical Model

Tapas Dey*[1], Anjali Bansal[2] and Laxmi Kant Dwivedi[3]

Abstract: Population projection based on the cohort-component method fails to incorporate uncertainty component. So, to overcome this issue, we have used the probabilistic approach using Monte Carlo Markov Chains (MCMC) method. We obtained data from previous censuses from 1901 to 2011, and applied the Bayes rule to get the posterior distributions. A four-parameter logistic growth model was used to project the population of eight metropolitan cities in India by taking the sample nodes for 2021-2071. The results observed smooth curves of the posterior density of the nodes, and the curves are bell-shaped, which indicates asymptotically normal. The four-parameter logistic model shows a closed fit to observed data for the years 1901 to 2011. The dotted line provides the 95% highest posterior density region, which is not large for the different population estimates. The projected population of Mumbai, Delhi, Kolkata, Nagpur, Indore and Meerut will be stagnant after 2051 and Chennai and Hyderabad will be stabilized after 2061. The study shows that the logistic growth model using the MCMC technique is suitable for the population projection, and holds a significant importance in the absence of latest Census of India 2021. The projection of these metro cities will help to formulate the future strategies that need to be changed with population growth. The approach can be used as an extension to the classical approach to predict future population as it yields more accurate estimates to measure uncertainty.

Keywords: Monte Carlo Markov Chains, Bayesian Hierarchical Model, Population Projection, Logistic Model, Metro City, WinBUGS.

Introduction

Population projections are always requisite for social, economic and infrastructure planning. They are used mainly by the government and private sectors for marketing decisions, formulating strategies and also provides inputs to social and health research (Raftery, Li, Ševčíková, Gerland, and Heilig, 2012; Verma, Singh, Pundir, and Singh, 2017). The population size directly affects the development of nation, so it is essential to project the future population. There are mainly two types of approaches in statistics to study the population projection; the first is the frequentist or conventional approach and the second approach is Bayesian. Demographers have used various methods to project the population. National and international organizations have widely adopted the cohort component model devised by Whelpton in 1936 (Whelpton 1936; UN 2004; GAD 2004; Burch 2018). The United Nations Population Fund also projects population based on the cohort component method (UNFPA 2020). However, the traditional demographic model does not fully assess uncertainties about the quantities of the future population. Also, incorporating information from different sources requires adjustment to maintain consistency (Alexander and Alkema, 2021). The methods widely used in the demographic literature are linear,

* Corresponding Author
[1] Research Scholar, International Institute for Population Sciences, Mumbai. Email: tapashb06@gmail.com
[2] Research Scholar, International Institute for Population Sciences, Mumbai. Email: anjali.bansal35@gmail.com
[3] Professor, Department of Survey Research and Data Analytics, International Institute for Population Sciences, Mumbai. Email: laxmikant@iipsindia.ac.in

exponential, logistic, Cohort-component, and ARIMA time series model (Stevenson 2007; Smith 1987; Burch 2018). The MCMC method in our study is simple and transparent and tackles uncertainties (Raftery et al., 2012; Singh, Pandey, and Rahul, 2007). Uncertainties are an issue in forecasting the population, which needs to be tackled to provide unbiased estimates for future planning.

The probabilistic population projection provides a comprehensive description of the uncertainty range (Raftery et al. 2012; Gneiting 2008; Lutz and Schröder 2010; Alkema et al. 2015). The main advantage of the following method is that we can get a clear scenario of the future population with the confidence region in a probabilistic approach which is commonly termed as Highest Posterior Density (HPD) region. There is a 95% chance or probability that the possible estimate values of parameters lie within the interval (Hespanhol et al.,2018). So, in this study, we developed a probabilistic Bayesian approach to estimate the city level projection. The probabilistic approach considers the full range of future possibilities, which is driven to assess the reality of changes in population forecasts and estimates (Raftery and Ševčíková, 2021). The approach reasonably fit well when the proper data is not available to provide the future and past population.

Today, more than half of the world's population lives in urban areas, and this figure is expected to increase, with the majority increase occurring in Asia and Africa (UN DESA 2014). India has been a major supplier to the increased urban population, attributing to her tremendous demographic pressure and her dynamics of urbanization (Haque and Patel, 2018). Compared to other Asian and developing countries, more than 70% of the urbanities live in cities today (Kundu 2014), which is also expected to increase over time. Therefore, it is essential to estimate the future population of these cities, as they have significant importance in the economic, health and demographic situation within a country. The projections will also help formulate the future strategies that need to be changed with population growth. So, the present paper projects the population of the cities namely Mumbai, Delhi, Kolkata, Chennai, Hyderabad, Nagpur, Indore and Meerut up to the year 2071.

Materials and Methods

Data

The city population for Mumbai, Delhi, Kolkata, Chennai, Hyderabad, Nagpur, Indore and Meerut were used, which was derived from the previous Censuses, 1901 to 2011. The Urban Agglomerations (U.A.) population data is obtained from the Census. According to the Census table, these are the Class I cities (Population 100,000 and above).

Methodology

Bayesian analysis was employed to project the population. At first, a probability model was formulated for the analysis, a prior distribution was decided, and construction of a likelihood was based on the collected data and the model. The likelihood was combined with the prior distribution. After that, posterior probability density function (pdf) was obtained. The future estimates were carried out based on it.

We have assumed that, Y_i denotes the city population's size for any city for the year t_i (i = 1,2, . . ., 11), i refers to successive censuses starting from 1901 for which i = 1. We have used a four-parameter logistic model. Let the general regression equation be $\mathrm{Y_i} = \mu_\mathrm{i} + \varepsilon_\mathrm{i}$

Where, μ_i is the deterministic part and ε_i is the disturbance part. It is supposed to follow independently identically distributed normal variable (i.i.d.-normal) with mean 0 and precision (=1/Variance) tau(τ). Our four-parameter logistic model is:

$$\mu_i = \frac{\theta_1\theta_2}{\theta_1 + (\theta_2 - \theta_1)\ e^{\theta_3(t_i - mean(t[\quad]))/sd(t[\quad])}} + \theta_4$$

Some re-parametrization is made to run smoothly in the WinBUGS.

$\theta_1 = e^{\emptyset_1}, \theta_2 = \emptyset_2, \theta_3 = \emptyset_3, \theta_4 = e^{\emptyset_4}$

Now, we have to provide prior distributions to all the parameters $\theta_1, \theta_2, \theta_3, \theta_4$ and τ. We don't have information about all the parameters. We provided non-informative priors to them, N (0,0.001). We provided a prior distribution to all the parameters $\emptyset_1, \emptyset_2, \emptyset_3, \emptyset_4$ present in the model. The inference was based on the data. The methodology about the choice of non-informative priors was in detail given by Spiegelhalter, Thomas, Best, and Gilks (Spiegelhalter et al. 1996). Here we assigned Normal (0, 0.001) (variance=1/0.001) before all of the parameters $\theta_1, \theta_2, \theta_3, \theta_4$ and Gamma (0.01, 0.01) prior to the parameter τ.

Markov Chain Monte Carlo (MCMC) method, a simulation technique, is used to predict population (Lee 1992) using freely available software WinBUGS, i.e., Bayesian inference using Gibbs Sampling for Windows. MCMC method is a computer-intensive method that can reduce the problems related to the analytical intractability of complex Bayesian models. The simulation process in MCMC provides a suitable way to obtain the probability distributions and provides probability intervals for the required model estimates for forecasting. The process error and the parameter error are allowed in this probabilistic approach. The MCMC method is technically demanding as this method deals with the statistical complex models. MCMC generates large independent iterated samples using the distribution of random numbers, and it takes the average function value from it. There are various diagnostics tools in WinBUGS. In the MCMC method, the analysis provides reliable estimates after running the chains sufficiently by many iterations. We run chains of each parameter for a long time. However, it was tough to detect the point of convergence conclusively when running the chains or simulations. So, various tools were developed to diagnose the situation whether the chain has converged or not. In the MCMC method, the next samples depend on the preceding, or existing samples termed a Markov chain. In WinBUGS, we have set two chains in the model for our study. It is also a way to run multiple chains at a time to check the convergence of simulations. When the different chains do not sufficiently mix after a long run, it lacks convergence. After confirming the convergence of the chains through diagnostics, we have run the simulations to obtain further samples for posterior inference. The history plot depicts mixing chains of sample values versus iterations (Congdon, 2007; Gilks, 1996; Singh, Pandey, and Rahul, 2007).

In 1998, Neal suggested that the iteration can be done by thinning chains or over relaxation (Neal 1998). We discarded all but considered every k^{th} observation with a goal of reducing autocorrelation, which is termed as thinning of chains. Generating multiple samples at each iteration and considering one correlated negatively with the current value is commonly known as over relaxation. In the time of thinning, every k^{th} iteration was stored where k, the value of thin. Taking k>1 helps to reduce the autocorrelation. We have set the value of thin as 2, and over-relaxation was used here. There was an increase in simulation timings, but there was a reduction of within chain autocorrelation. As a result, there were fewer iterations were necessarily required. In 1998, Brooks and Gelman proposed Brooks-Gelman-Rubin (bgr) convergence statistic. The Gelman Rubin statistic is denoted by R, a popular diagnostic measure for the MCMC convergence (Brooks and Gelman 1998). Greenline shows the width of the central 80% interval of the pooled runs, the blue line shows the average width of the 80% intervals within the individual runs, and the red line represents its ratio R, which is (R= pooled/within). These are normalized for plotting. Brooks and Gelman (1998) suggested that we have to concern about the convergence of R to 1 and convergence of pooled and within interval widths.

Another tool is the trace plot of updated chains of parameters. If we have more samples, estimates will be more accurate. If we are not comfortable with the current trace plots, we will run further simulations to get more accurate samples for posterior inference. So, we can discard the previous samples and can consider the current ones for future estimates. We can go further based on summary statistics obtained from WinBUGS.

Results

In the MCMC method, the chain gradually forgets its initial state and eventually reaches invariant or stationary distribution. Usually, burn-in samples are discarded for this kind of calculation (Gilks, 1996). After discarding 10,000 initial iterations, we studied the parameters based on later 50,000 iterations. During the following updates, none of the diagnostics indicated any of the symptoms of non-convergence of chains. The following number of iterations is required to run after the convergence of chains is assessed depending on each parameter's Monte Carlo (MC) error. MC error is obtained by the difference between the mean of sampled values used here to estimate the posterior mean for each parameter and the true mean for posterior distribution. It is suggested by the WinBUGS manual that we should run the simulation process until MC error reaches less than 5 per cent of sample standard deviation (SD). Here we observed all parameters of our interest. We obtained the value of R is close to 1 from the Brooks-Gelman-Rubin (bgr) diagnostic for all the nodes $\theta_1, \theta_2, \theta_3, \theta_4$.

We have used a four-parameter logistic growth model in our study. $\theta_1, \theta_2, \theta_3, \theta_4$ −These are the parameters in our model, where the upper asymptote of the logistic curve approaches to θ_2+θ_4, which is the population's carrying capacity. Here, -θ_3 represents the growth of the population. Here $\phi_1, \phi_2, \phi_3, \phi_4$ are the parameters which are used in re-parametrization to run the model smoothly in the WinBUGS. Since we don't have much knowledge about the parameters, we have provided non-informative priors. We have assigned Normal (0, 0.001) to all the parameters $\theta_1, \theta_2, \theta_3, \theta_4$.The summary statistics for all the nodes are provided in Tables 1-8. It is also observed that the MC error for all nodes is less than their sample standard deviation. Summary statistics contain each node's mean, median, MC error, and HPD region. Using the summary statistics tables, we depicted the city population on the basis of the assumed model. We also provided the projected tables for city population with graphical representations for all the metropolitan cities of India.

Table 1: Summary statistics of the model for Mumbai

				HPD Region		
Node	**Mean**	**SD**	**MC error**	**2.50%**	**Median**	**97.50%**
phi[1]	2.722	0.02083	9.04E-04	2.68	2.722	2.757
phi[2]	26.03	3.917	0.1753	20.36	25.63	35.67
phi[3]	-2.744	0.4118	0.01977	-3.643	-2.619	-2.104
phi[4]	-21.13	22.59	1.04	-62.98	-11.81	-0.00964
sigma	0.381	0.1278	0.004041	0.1912	0.366	0.6782
theta[1]	15.21	0.3164	0.01373	14.59	15.21	15.75
theta[2]	26.03	3.917	0.1753	20.36	25.63	35.67
theta[3]	-2.744	0.4118	0.01977	-3.643	-2.619	-2.104
theta[4]	0.2907	0.3604	0.01763	4.44E-28	7.42E-06	0.9904

Table 2: Summary statistics of the model for Delhi

				HPD Region		
Node	**Mean**	**SD**	**MC error**	**2.50%**	**Median**	**97.50%**
phi[1]	2.487	0.01459	6.92E-04	2.454	2.485	2.51
phi[2]	29.21	3.386	0.1443	24.32	28.62	37.57
phi[3]	-2.939	0.1748	0.008082	-3.301	-2.937	-2.588
phi[4]	-31.1	19.8	0.7538	-66.47	-28.36	-1.697
sigma	0.2249	0.05955	0.001113	0.1418	0.2143	0.3698
theta[1]	12.03	0.1751	0.008304	11.64	12	12.31
theta[2]	29.21	3.386	0.1443	24.32	28.62	37.57
theta[3]	-2.939	0.1748	0.008082	-3.301	-2.937	-2.588
theta[4]	0.01164	0.05524	0.002454	1.35E-29	4.83E-13	0.1832

Table 3: Summary statistics of the model for Kolkata

				HPD Region		
Node	**Mean**	**SD**	**MC error**	**2.50%**	**Median**	**97.50%**
phi[1]	2.496	0.04726	0.002299	2.417	2.515	2.556
phi[2]	19.01	4.268	0.1929	13.9	18.91	27.75
phi[3]	-2.166	0.4797	0.02362	-3.151	-1.989	-1.512
phi[4]	-14.23	17.63	0.8068	-56.55	-5.926	0.4353
sigma	0.35	0.1053	0.002616	0.2087	0.3299	0.6084
theta[1]	12.15	0.5684	0.02759	11.22	12.36	12.89
theta[2]	19.01	4.268	0.1929	13.9	18.91	27.75
theta[3]	-2.166	0.4797	0.02362	-3.151	-1.989	-1.512
theta[4]	0.4901	0.6025	0.0298	2.75E-25	0.002668	1.545

Table 4: Summary statistics of the model for Chennai

				HPD Region		
Node	**Mean**	**SD**	**MC error**	**2.50%**	**Median**	**97.50%**
phi[1]	1.891	0.02023	7.38E-04	1.854	1.892	1.933
phi[2]	22.17	9.932	0.3607	11.33	19.39	48.84
phi[3]	-2.003	0.3082	0.0145	-2.848	-1.93	-1.589
phi[4]	-16.56	16.98	0.552	-61.45	-11.58	-0.8842
sigma	0.217	0.05725	7.83E-04	0.1359	0.207	0.3557
theta[1]	6.625	0.1337	0.004873	6.388	6.631	6.909
theta[2]	22.17	9.932	0.3607	11.33	19.39	48.84
theta[3]	-2.003	0.3082	0.0145	-2.848	-1.93	-1.589
theta[4]	0.06585	0.1214	0.005724	2.06E-27	9.31E-06	0.4131

Table 5: Summary statistics of the model for Hyderabad

				HPD Region		
Node	**Mean**	**SD**	**MC error**	**2.50%**	**Median**	**97.50%**
phi[1]	1.649	0.02897	0.001141	1.583	1.647	1.7
phi[2]	18.43	10.04	0.3894	10.19	15.34	46.56
phi[3]	-2.789	0.4213	0.01686	-3.622	-2.785	-2.013
phi[4]	-1.331	1.168	0.04425	-5.315	-1.062	-0.5759
sigma	0.1739	0.05036	9.42E-04	0.1049	0.1646	0.2977
theta[1]	5.204	0.1503	0.005914	4.87	5.193	5.471
theta[2]	18.43	10.04	0.3894	10.19	15.34	46.56
theta[3]	-2.789	0.4213	0.01686	-3.622	-2.785	-2.013
theta[4]	0.3349	0.1317	0.004859	0.004918	0.3457	0.5622

Table 6: Summary statistics of the model for Nagpur

				HPD Region		
Node	**Mean**	**SD**	**MC error**	**2.50%**	**Median**	**97.50%**
phi[1]	0.6986	0.02253	8.33E-04	0.6371	0.702	0.7344
phi[2]	5.554	5.799	0.2616	3.106	4.443	13.45
phi[3]	-2.2	0.3063	0.01325	-3.079	-2.18	-1.656
phi[4]	-31.05	19.94	0.7255	-66.04	-31.29	-2.274
sigma	0.06536	0.01863	3.89E-04	0.04035	0.06168	0.1115
theta[1]	2.011	0.04494	0.001657	1.891	2.018	2.084
theta[2]	5.554	5.799	0.2616	3.106	4.443	13.45
theta[3]	-2.2	0.3063	0.01325	-3.079	-2.18	-1.656
theta[4]	0.008766	0.02745	0.001242	2.08E-29	2.59E-14	0.1029

Table 7: Summary statistics of the model for Indore

				HPD Region		
Node	**Mean**	**SD**	**MC error**	**2.50%**	**Median**	**97.50%**
phi[1]	2.709	0.02188	5.47E-04	2.659	2.711	2.747
phi[2]	78.47	17.97	0.2801	47.16	77.19	117.4
phi[3]	-2.102	0.144	0.003496	-2.457	-2.074	-1.903
phi[4]	-18.68	18.72	0.5142	-62.9	-13.54	-0.284
sigma	0.3887	0.1186	0.001156	0.2266	0.3663	0.6818
theta[1]	15.01	0.3264	0.008105	14.28	15.05	15.6
theta[2]	78.47	17.97	0.2801	47.16	77.19	117.4
theta[3]	-2.102	0.144	0.003496	-2.457	-2.074	-1.903
theta[4]	0.1227	0.223	0.007261	4.803E-28	1.32E-06	0.7528

Table 8: Summary statistics of the model for Meerut

				HPD Region		
Node	**Mean**	**SD**	**MC error**	**2.50%**	**Median**	**97.50%**
phi[1]	2.289	0.03777	0.001329	2.232	2.282	2.394
phi[2]	23.63	11.06	0.3946	15.64	19.79	59.63
phi[3]	-3.215	0.6293	0.02305	-4.249	-3.312	-1.799
phi[4]	-1.345	5.562	0.2227	-20.97	0.2188	0.473
sigma	0.3199	0.1424	0.00448	0.1712	0.278	0.7194
theta[1]	9.875	0.3827	0.01356	9.314	9.799	10.96
theta[2]	23.63	11.06	0.3946	15.64	19.79	59.63
theta[3]	-3.215	0.6293	0.02305	-4.249	-3.312	-1.799
theta[4]	1.121	0.4232	0.0163	7.81E-10	1.245	1.605

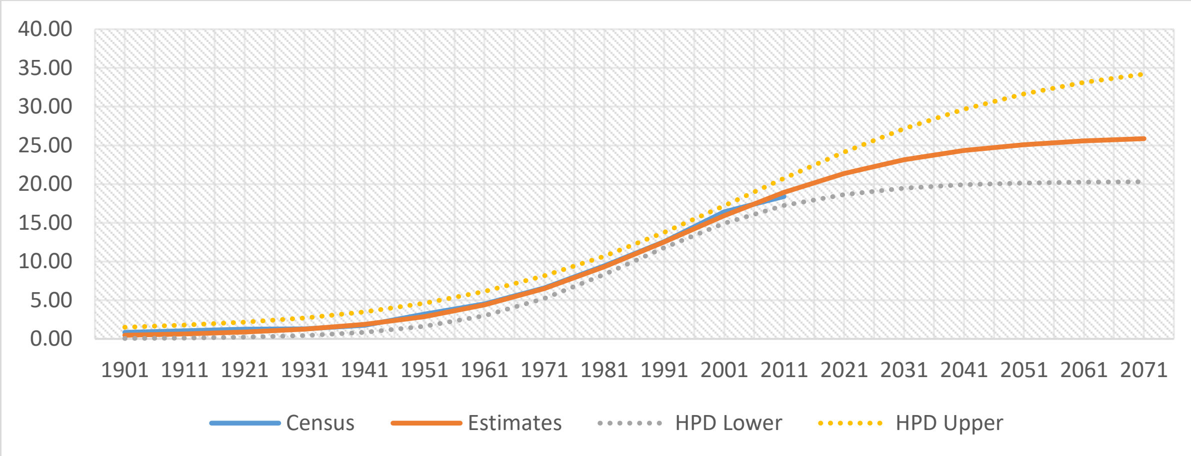
Table 9: Projected Population and HPD Region for Mumbai

Year		**Population Estimates**		**HPD Region**	
	Census (in Million)	**Estimated (Mean)**	**Estimated (Median)**	**2.50%**	**97.50%**
1901	0.84	0.50	0.27	0.06	1.53
1911	1.05	0.65	0.45	0.11	1.79
1921	1.29	0.88	0.72	0.22	2.16
1931	1.32	1.27	1.17	0.43	2.71
1941	1.75	1.90	1.87	0.85	3.50
1951	3.22	2.89	2.93	1.63	4.61
1961	4.52	4.39	4.49	3.01	6.13
1971	6.60	6.52	6.64	5.24	8.14
1981	9.42	9.30	9.35	8.32	10.70
1991	12.60	12.56	12.45	11.81	13.76
2001	16.43	15.91	15.59	14.94	17.18
2011	18.39	18.94	18.42	17.23	20.73
2021		21.37	20.70	18.66	24.13
2031		23.13	22.39	19.47	27.16
2041		24.33	23.55	19.91	29.68
2051		25.10	24.33	20.13	31.65
2061		25.58	24.82	20.24	33.14
2071		25.88	25.13	20.30	34.22

Mumbai

Mumbai is the highest populated metropolitan city in India. From the summary statistics tables, the upper asymptote or the carrying capacity of Mumbai is estimated as ($\theta_2+\theta_4$= 26.03+0.29) 26.32 million (Table 1). Using the MCMC tools, it is clear that the data fits the following 4-parameter logistic model (Table 9). There was a lot of problem in the data before independence. After independence, the data was well fitted. According to the census, the population of Mumbai was 4.52, 6.60, 9.42, 12.60, 16.43, 18.39 million respectively in the years 1961, 1971, 1981, 1991, 2001 and 2011. And the fitted population for Mumbai was 4.39, 6.52, 9.30, 12.56, 15.91, 18.94 in 1961, 1971, 1981, 1991, 2001 and 2011, respectively. The projected population will reach to the level of 21.37 million for the year 2021. In 2051, 2061 and 2071, the city population may reach 25.10, 25.58 and 25.88 million respectively. It seems that the city population will be stagnant after 2051. The upper and lower estimates are also given in the tables. The graphical representation of the models fitted and projected for the population shows in figure 1. For Mumbai, the differences between the census value and the estimated value were less than 3% from the year 1961 except for the year 2001. The difference was 3.2% for the year 2001 (Figure 1).

Figure 1: Population Projection for Mumbai (1901-2071)



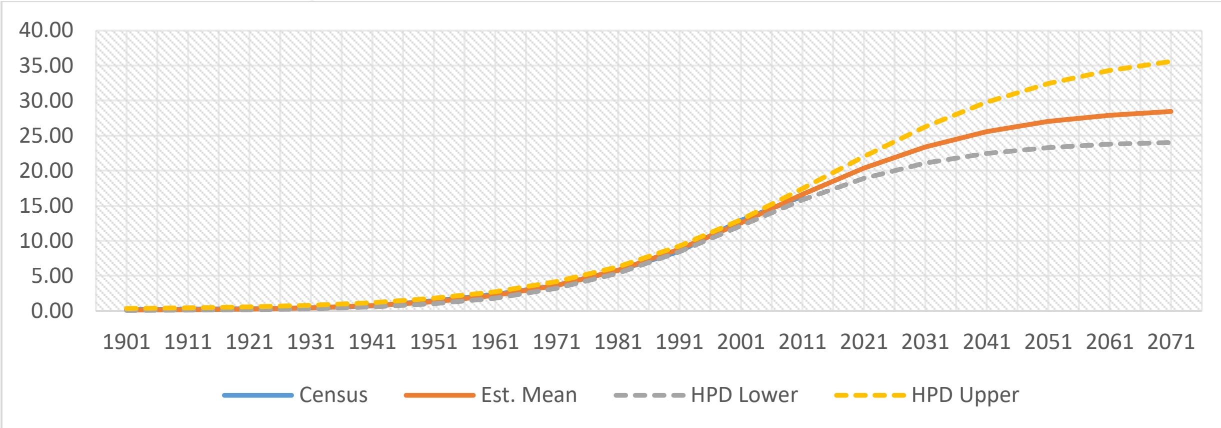
Delhi

Delhi is the second most populated metropolitan city in India. The model fits the data for Delhi very well (Table 10). From the summary statistics tables, the upper asymptote or the carrying capacity of Delhi is estimated as ($\theta_2+\theta_4$= 29.21+0.01) 29.22 million (Table 2). The projected population for Delhi will be 20.35 million for the year 2021. The Delhi population may reach to 27.01, 27.and 28.46 million for the years 2051, 2061 and 2071 respectively. The projected population for Delhi indicates that the population will be stagnant after 2051. For Delhi, the differences between the census value and the estimated value were less than 3% from the year 1961 except for the years 1961 and 1991. The differences were 7.2% and 3.8% for the years 1961 and 1991 respectively (Figure2).

Table 10: Projected Population and HPD Region for Delhi

Year	Census (in Million)	Population Estimates: Estimated (Mean)	Population Estimates: Estimated (Median)	HPD Region: 2.50%	HPD Region: 97.50%
1901	0.21	0.09	0.08	0.05	0.33
1911	0.24	0.16	0.15	0.09	0.42
1921	0.30	0.26	0.25	0.16	0.56
1931	0.45	0.45	0.44	0.30	0.80
1941	0.70	0.76	0.76	0.55	1.18
1951	1.44	1.30	1.30	1.01	1.78
1961	2.36	2.19	2.20	1.82	2.73
1971	3.65	3.62	3.63	3.20	4.17
1981	5.76	5.78	5.79	5.37	6.29
1991	8.47	8.79	8.79	8.43	9.23
2001	12.90	12.54	12.49	12.12	13.01
2011	16.35	16.59	16.45	15.81	17.41
2021		20.35	20.11	18.89	22.00
2031		23.38	23.03	21.08	26.24
2041		25.57	25.13	22.47	29.76
2051		27.01	26.51	23.30	32.42
2061		27.91	27.37	23.76	34.30
2071		28.46	27.89	24.02	35.56

Figure 2: Population Projection for Delhi (1901-2071)



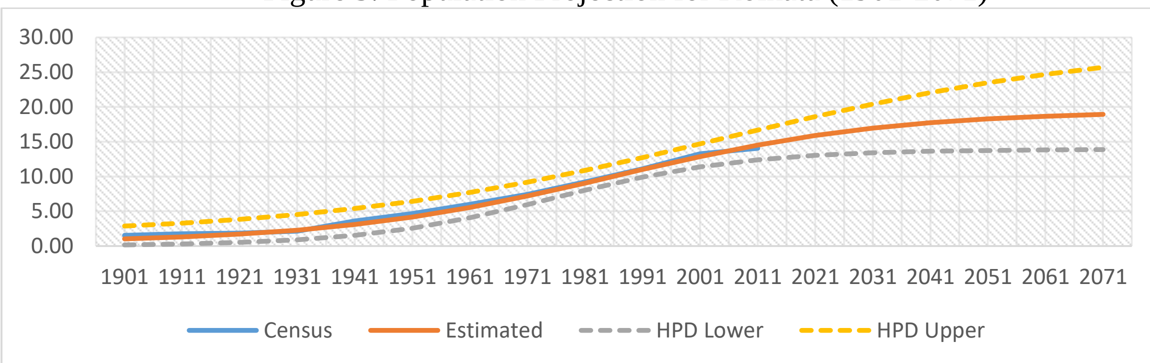
Kolkata

The upper asymptote or the carrying capacity of Kolkata is estimated as (θ_2+θ_4= 19.01+0.49) 19.50 million (Table 3). According to the census, the population of Kolkata is 6.01, 7.45, 9.23, 11.11, 13.25, 14.06 million, respectively, in the years 1961, 1971, 1981, 1991, 2001 and 2011. And the derived fitted population for Kolkata is 5.52, 7.18, 9.05, 11.00, 12.86 and 14.51 million in 1961, 1971, 1981, 1991, 2001 and 2011, respectively. In the year 2021, the projected population of Kolkata will reach to the level of 15.87 million. The city population may reach to 18.28, 18.67 and 18.94 million in 2051, 2061 and 2071, respectively (Table 11). It seems that population of Kolkata will be stagnant after 2051. For Kolkata, the differences between the census value and the estimated value were less than 3% from the year 1961 except for the years 1961, 1971 and 2011. The differences were 8.2%, 3.6% and 3.2% for the years 1961, 1971 and 2011, respectively. (Figure 3).

Table 11: Projected Population and HPD Region for Kolkata

Year		**Population Estimates**		**HPD Region**	
	Census (in Million)	**Estimated (Mean)**	**Estimated (Median)**	**2.50%**	**97.50%**
1901	1.52	1.05	0.82	0.16	2.89
1911	1.76	1.33	1.18	0.28	3.31
1921	1.87	1.73	1.67	0.50	3.85
1931	2.12	2.29	2.34	0.89	4.53
1941	3.60	3.09	3.22	1.54	5.39
1951	4.69	4.15	4.36	2.57	6.44
1961	6.01	5.52	5.75	4.06	7.70
1971	7.45	7.18	7.36	5.96	9.19
1981	9.23	9.05	9.11	8.02	10.87
1991	11.11	11.00	10.89	9.91	12.73
2001	13.25	12.86	12.56	11.38	14.68
2011	14.06	14.51	14.05	12.39	16.67
2021		15.87	15.28	13.03	18.61
2031		16.93	16.26	13.41	20.42
2041		17.72	17.01	13.62	22.06
2051		18.28	17.57	13.75	23.49
2061		18.67	17.97	13.82	24.70
2071		18.94	18.26	13.85	25.69

Figure 3: Population Projection for Kolkata (1901-2071)



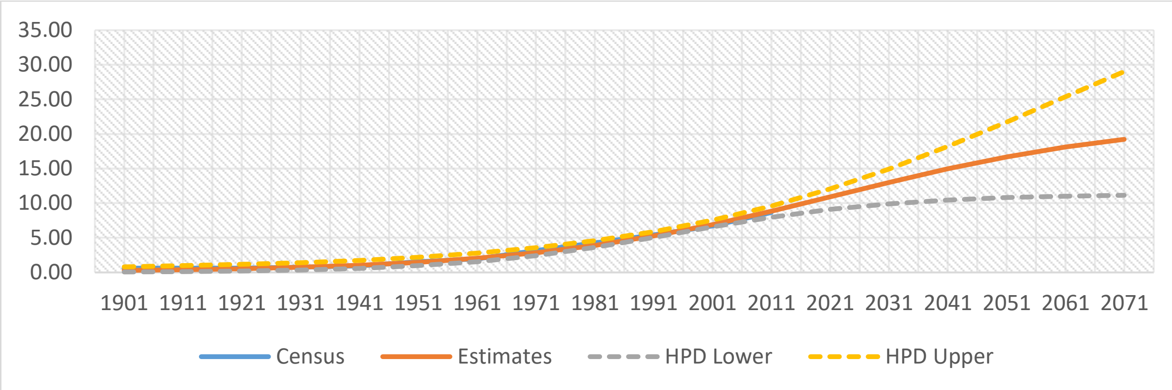
Chennai

From the summary statistics tables, the carrying capacity of Chennai is estimated as (θ_2+θ_4= 22.17+0.07) 22.24 million (Table 4). According to the census, the population of Chennai is 1.94, 3.16, 4.27, 5.42, 6.69, 8.65 million, respectively, in 1961, 1971, 1981, 1991, 2001 and 2011. And the fitted population for Chennai is 2.04, 2.84, 3.90, 5.25, 6.91 and 8.83 million in 1961, 1971, 1981, 1991, 2001 and 2011, respectively. The fitted population for Chennai will be 10.90 million for the year 2021. The city population may reach to 18.11 and 19.24 million in 2061 and 2071, respectively (Table 12). The projected Chennai population shows that it will be stagnant after 2061. For Chennai, the differences between the census value and the estimated value were less than 4% from the year 1961 except for the years 1961, 1971 and 1981. The differences were 5.2%, 10.1% and 8.7% for the years 1961, 1971 and 1981, respectively. (Figure 4).

Table 12: Projected Population and HPD Region for Chennai

Year		**Population Estimates**		**HPD Region**	
	Census (in Million)	**Estimated (Mean)**	**Estimated (Median)**	**2.50%**	**97.50%**
1901	0.59	0.28	0.27	0.07	0.82
1911	0.60	0.38	0.38	0.12	0.96
1921	0.63	0.53	0.55	0.20	1.15
1931	0.77	0.74	0.78	0.35	1.40
1941	0.93	1.03	1.10	0.58	1.74
1951	1.54	1.45	1.55	0.96	2.19
1961	1.94	2.04	2.16	1.55	2.79
1971	3.16	2.84	2.98	2.43	3.57
1981	4.27	3.90	4.02	3.62	4.59
1991	5.42	5.25	5.31	5.05	5.90
2001	6.69	6.91	6.83	6.58	7.55
2011	8.65	8.83	8.53	7.97	9.59
2021		10.90	10.30	9.10	12.05
2031		12.99	12.03	9.91	14.93
2041		14.96	13.61	10.46	18.18
2051		16.68	14.98	10.81	21.71
2061		18.11	16.11	11.02	25.38
2071		19.24	16.99	11.15	29.02

Figure 4: Population Projection for Chennai (1901-2071)



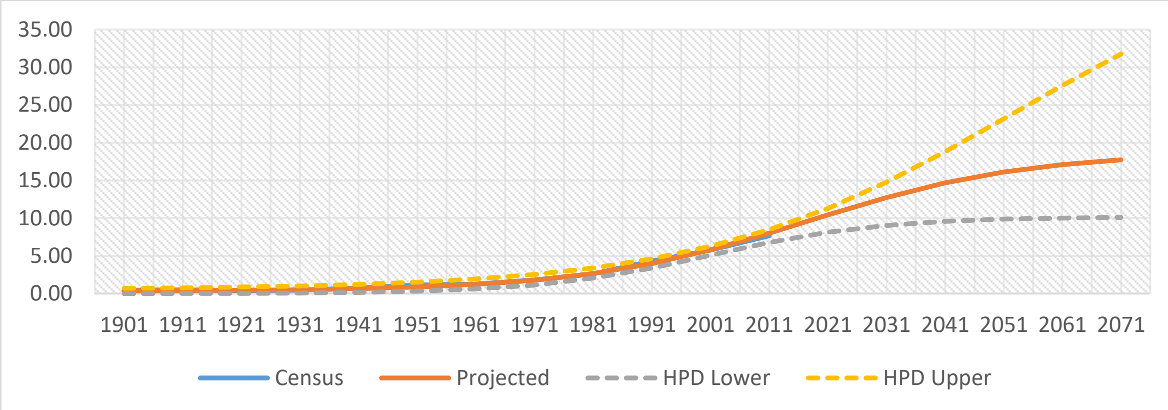
Hyderabad

The carrying capacity of Hyderabad is estimated as (θ_2+θ_4= 18.43+0.33) 18.76 million (Table 5). The fitted population for Hyderabad is 1.22, 1.79, 2.68, 3.99, 5.79 and 8.01 million in 1961, 1971, 1981, 1991, 2001 and 2011, respectively. The projected population for Hyderabad will be 10.43 for the year 2021. The city population may reach 17.10 and 17.75 million in 2061 and 2071 respectively. (Table 13). The projection shows that the city will attain stagnant population after 2061 (Figure 5).

Table 13: Projected Population and HPD Region for Hyderabad

Year		Population Estimates		HPD Region	
	Census (in Million)	Estimated (Mean)	Estimated (Median)	2.50%	97.50%
1901	0.45	0.37	0.39	0.02	0.70
1911	0.50	0.40	0.42	0.03	0.77
1921	0.41	0.45	0.47	0.05	0.87
1931	0.47	0.52	0.55	0.09	1.01
1941	0.74	0.65	0.69	0.17	1.21
1951	1.14	0.87	0.92	0.32	1.50
1961	1.25	1.22	1.29	0.62	1.93
1971	1.82	1.79	1.89	1.15	2.54
1981	2.61	2.68	2.79	2.06	3.41
1991	4.34	3.99	4.08	3.41	4.61
2001	5.76	5.79	5.77	5.09	6.27
2011	7.68	8.01	7.73	6.78	8.47
2021		10.43	9.73	8.13	11.31
2031		12.74	11.51	9.04	14.78
2041		14.66	12.91	9.58	18.80
2051		16.11	13.92	9.88	23.17
2061		17.10	14.59	10.03	27.59
2071		17.75	15.02	10.11	31.74

Figure 5: Population Projection for Hyderabad (1901-2071)



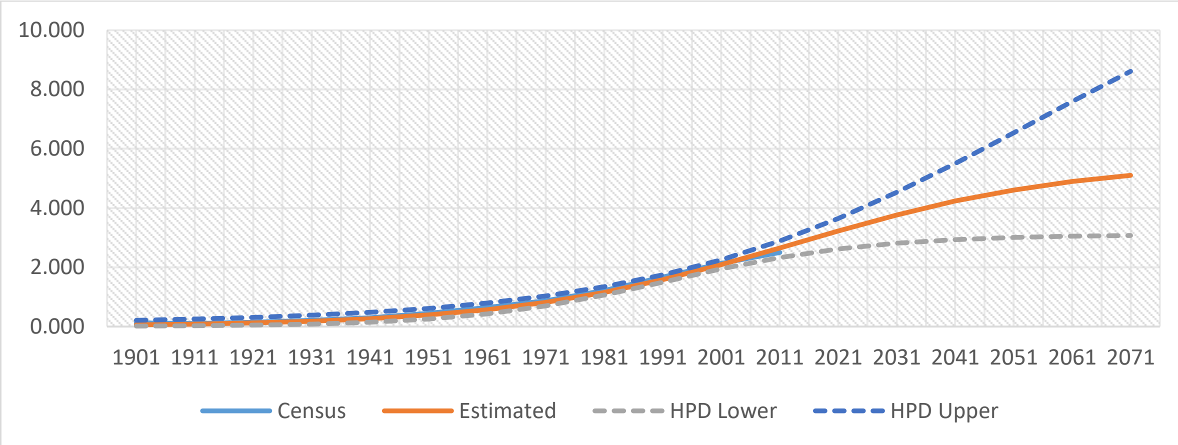
Nagpur

The carrying capacity of Nagpur is estimated as ($\theta_2+\theta_4$= 5.55+0.01) 5.56 million (Table 6). The fitted population for Nagpur is 0.57, 0.82, 1.15, 1.58, 2.09 and 2.65 million in 1961, 1971, 1981, 1991, 2001 and 2011, respectively. In the year 2021, the Nagpur will reach to the level of 3.23 million population. The city population may reach 4.61, 4.90 and 5.10 million in 2051, 2061 and 2071, respectively (Table 14). It shows that the population will be stagnant after 2051 (Figure 6).

Table 14: Projected Population and HPD Region for Nagpur

Year	**Population Estimates**			**HPD Region**	
	Census (in Million)	**Estimated (Mean)**	**Estimated (Median)**	**2.50%**	**97.50%**
1901	0.13	0.06	0.06	0.01	0.21
1911	0.10	0.09	0.09	0.03	0.25
1921	0.15	0.12	0.14	0.05	0.31
1931	0.22	0.18	0.20	0.08	0.38
1941	0.30	0.27	0.30	0.15	0.48
1951	0.45	0.39	0.44	0.26	0.62
1961	0.64	0.57	0.64	0.43	0.79
1971	0.87	0.82	0.90	0.70	1.03
1981	1.22	1.15	1.23	1.06	1.34
1991	1.66	1.58	1.63	1.50	1.75
2001	2.13	2.09	2.08	1.94	2.26
2011	2.50	2.65	2.53	2.33	2.89
2021		3.23	2.96	2.62	3.65
2031		3.77	3.34	2.81	4.52
2041		4.23	3.65	2.94	5.50
2051		4.61	3.88	3.01	6.54
2061		4.90	4.06	3.05	7.60
2071		5.10	4.18	3.08	8.61

Figure 6: Population Projection for Nagpur (1901-2071)



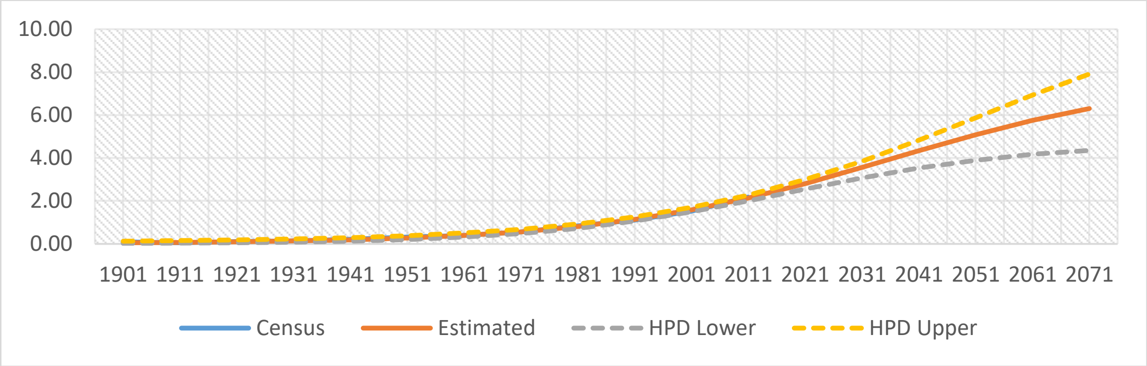
Indore

The carrying capacity of Indore is estimated as (θ_2+θ_4= 78.47+0.12) 78.59 million (Table 7) According to the census, the population of Indore is 0.39, 0.56, 0.83, 1.11, 1.52, 2.17 million respectively in the years 1961, 1971, 1981, 1991, 2001 and 2011. And the derived fitted population for Indore is 0.39, 0.56, 0.80, 1.13, 1.57 and 2.13 million in the year 1961, 1971, 1981, 1991, 2001 and 2011 respectively. The Indore population will be 2.80 million for the year 2021. The city population may reach to 5.09, 5.75 and 6.31 million for the years 2051, 2061 and 2071, respectively (Table 15). The fitted population of Indore depicts that the Indore population will be stagnant after 2051. For Indore, the differences between the census value and the estimated value were around or less than 3% from the year 1961. (Figure 7).

Table 15: Projected Population and HPD Region for Indore

Year		Population Estimates		HPD Region	
	Census (in Million)	Estimated (Mean)	Estimated (Median)	2.50%	97.50%
1901	0.10	0.05	0.04	0.02	0.13
1911	0.05	0.07	0.06	0.03	0.15
1921	0.11	0.09	0.08	0.05	0.18
1931	0.14	0.13	0.12	0.08	0.22
1941	0.20	0.19	0.18	0.13	0.29
1951	0.31	0.27	0.26	0.20	0.38
1961	0.39	0.39	0.39	0.31	0.50
1971	0.56	0.56	0.56	0.48	0.68
1981	0.83	0.80	0.80	0.72	0.92
1991	1.11	1.13	1.13	1.06	1.25
2001	1.52	1.57	1.57	1.49	1.70
2011	2.17	2.13	2.11	2.00	2.27
2021		2.80	2.77	2.54	2.99
2031		3.56	3.50	3.07	3.85
2041		4.34	4.26	3.53	4.83
2051		5.09	4.98	3.89	5.87
2061		5.75	5.63	4.17	6.93
2071		6.31	6.18	4.35	7.92

Figure 7: Population Projection for Indore (1901-2071)



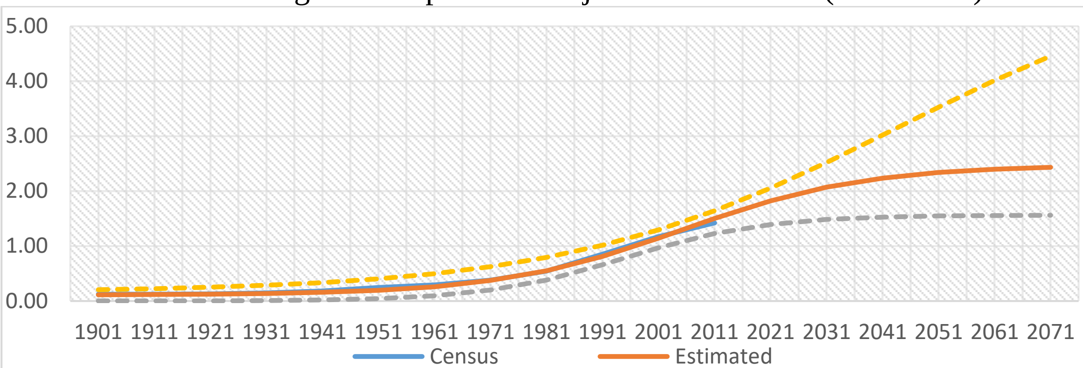
Meerut

The carrying capacity of Meerut is estimated as (θ_2+θ_4= 23.63+1.12) 24.75 million (Table 8). The population of Meerut is 0.29, 0.38, 0.54, 0.85, 1.17, 1.42 million in the years 1961, 1971, 1981, 1991, 2001 and 2011 respectively according to the census. And the fitted population for Meerut is 0.26, 0.37, 0.55, 0.81, 1.14 and 1.50 million in 1961, 1971, 1981, 1991, 2001 and 2011, respectively. The projected population for Meerut will reach to the level of 1.82 million for the year 2021. The city population of Meerut may reach to 2.34, 2.40 and 2.43 million in the year 2051, 2061 and 2071, respectively (Table 16). The projected population for the Meerut city shows that the city population will be stagnant after 2051 (Figure 8).

Table 16: Projected Population and HPD Region for Meerut

Year		**Population Estimates**		**HPD Region**	
	Census (in Million)	**Estimated (Mean)**	**Estimated (Median)**	**2.50%**	**97.50%**
1901	0.13	0.12	0.13	0.001	0.21
1911	0.12	0.12	0.13	0.002	0.22
1921	0.13	0.13	0.14	0.004	0.25
1931	0.15	0.14	0.15	0.009	0.29
1941	0.19	0.16	0.17	0.020	0.34
1951	0.25	0.20	0.21	0.044	0.40
1961	0.29	0.26	0.28	0.095	0.50
1971	0.38	0.37	0.40	0.198	0.62
1981	0.54	0.55	0.58	0.383	0.79
1991	0.85	0.81	0.84	0.658	1.01
2001	1.17	1.14	1.14	0.969	1.30
2011	1.42	1.50	1.44	1.228	1.64
2021		1.82	1.69	1.394	2.05
2031		2.07	1.86	1.483	2.52
2041		2.24	1.96	1.527	3.02
2051		2.34	2.03	1.547	3.53
2061		2.40	2.06	1.556	4.01
2071		2.43	2.08	1.561	4.45

Figure 8: Population Projection for Meerut (1901-2071)



Discussion

The present study is perhaps the first attempt to project the city level projection using the Bayesian population projections. All the previous studies on the projection were mainly based on the deterministic type projections (Dyson 2004; Registrar General of India 2006). In this study, we have projected the population for the eight metropolitan cities of India Mumbai, Delhi, Kolkata, Chennai, Hyderabad, Nagpur, Indore and Meerut using the data from the previous censuses, from 1901 to 2011, and projected the population of 2061 and 2071. It was found from the analysis that projected figures are very close to the census estimates. The results provided the estimated population mean, population median and HPD regions. For big cities like Mumbai and Delhi, the carrying capacity is 26.32 and 29.22 million, respectively. The projected population for Mumbai was 25.88 million for the year 2071, while for Delhi, it was 28.46 million. The HPD region is a distinct approach, which summarizes the sample space for a given probability. It also presents the possible shortest interval for one-dimensional cases and the possible smallest region for two-dimensional cases. Hyndman argues that the HPD region is the most effective summary measure for forecasting other than any common region (Hyndman 1996, 1995). It can provide asymmetry and multilocularity due to its flexibility. The density for every point inside the probability interval or region is greater than those points which are outside the region. Any value of the parameter inside the region has a high probability density. It contains the most credible values of parameters. The region's length of population estimates is not too large; it is quite narrow. It is also important to measure the accuracy of the predictive model. It is termed a Quadratic Loss Function. The sample mean and sample median are equally important in the Bayesian study. Estimates for the sample mean and median have a great role in reducing the quadratic loss errors in Bayesian estimates (Lehmann and Casella, 2006).

The study holds significant importance in the absence of the latest Census of India 2021. Population projections allow us to have better estimates to project demographic flows in a precise way. The population projection of these metro cities of India can be used for our country's economic flow. These projections are helpful to formulate the future strategies that need to be changed with population growth to provide better health infrastructure to the people and further to improve per-capita income.

In this study, we have made several assumptions because of sparse data on migration for India. We have applied non-informative priors to overcome this, which doesn't provide substantial information to the posterior distribution (Verma et al., 2017). In this study, we have not analyzed the constancy and accuracy of the estimates using other studies and surveys. The study's main objective was to project the city level population projection under the Bayesian approach. Despite the limitations, the logistic growth model using MCMC fits well with the past census data and can be used for future sub-population projections.

References

Alexander, Monica, and Leontine Alkema, 2021, "A Bayesian Cohort Component Projection Model to Estimate Adult Populations at the Subnational Level in Data-Sparse Settings." *ArXiv Preprint ArXiv:2102.06121*.

Alkema, Leontine, Patrick Gerland, Adrian Raftery, and John Wilmoth, 2015, "The United Nations Probabilistic Population Projections: An Introduction to Demographic Forecasting with Uncertainty." *Foresight (Colchester, Vt.)* 2015 (37): 19.

Brooks, Stephen P, and Andrew Gelman, 1998, "General Methods for Monitoring Convergence of Iterative Simulations." *Journal of Computational and Graphical Statistics* 7 (4): 434–55.

Burch, Thomas K, 2018, "The Cohort-Component Population Projection: A Strange Attractor for Demographers." In *Model-Based Demography,* 135–51. Springer.

Congdon, Peter, 2007, *Bayesian Statistical Modelling*. Vol. 704. John Wiley & Sons.

Dyson, Tim, 2004, *India's Population-the Future*. Oxford University Press.

GAD, 2004, "Population Database." 2004.

Gilks, Walter R., 1996, "Full Conditional Distributions." *Markov Chain Monte Carlo in Practice*, 75–88.

Gneiting, Tilmann, 2008, "Probabilistic Forecasting." *Journal of the Royal Statistical Society. Series A (Statistics in Society),* 319–21.

Haque, Ismail, and Priyank Pravin Patel, 2018, "Growth of Metro Cities in India: Trends, Patterns and Determinants." *Urban Research & Practice* 11 (4): 338–77.

Hyndman, Rob J., 1995, "Highest-density Forecast Regions for Nonlinear and Non-normal Time Series Models." *Journal of Forecasting* 14 (5): 431–41.

Hyndman, Rob J., 1996, "Computing and Graphing Highest Density Regions." *The American Statistician* 50 (2): 120–26.

Kundu, Debolina, 2014, "Urban Development Programmes in India: A Critique of JNNURM." *Social Change* 44 (4): 615–32.

Lee, Ronald D., 1992, "Stochastic Demographic Forecasting." *International Journal of Forecasting* 8 (3): 315–27. https://doi.org/10.1016/0169-2070(92)90050-J.

Lehmann, Erich L, and George Casella. 2006. *Theory of Point Estimation*. Springer Science & Business Media.

Lutz, Carsten, and Lutz Schröder., 2010, "Probabilistic Description Logics for Subjective Uncertainty." In *Twelfth International Conference on the Principles of Knowledge Representation and Reasoning*.

Neal, Radford M., 1998, "Suppressing Random Walks in Markov Chain Monte Carlo Using Ordered Overrelaxation." In *Learning in Graphical Models*, 205–28. Springer.

Raftery, Adrian E., Nan Li, Hana Ševčíková, Patrick Gerland, and Gerhard K. Heilig, 2012, "Bayesian Probabilistic Population Projections for All Countries." *Proceedings of the*

National Academy of Sciences of the United States of America 109 (35): 13915–21. https://doi.org/10.1073/PNAS.1211452109/-/DCSUPPLEMENTAL.

Raftery, Adrian E, and Hana Ševčíkovám 2021, "Probabilistic Population Forecasting: Short to Very Long-Term." *International Journal of Forecasting*.

Registrar General of India, 2006, "census of india 2001 population projections for india and states 2001-2026 report of the technical group on population projections constituted by the national commission on population."

Singh, O P, G S Pandey, and Rahul, 2007, "Population Projection of India (2001–2051) Using MCMC Technique in Bayesian Inference." *American Journal of Mathematical and Management Sciences* 27 (1–2): 139–52.

Smith, Stanley K., 1987, "Tests of Forecast Accuracy and Bias for County Population Projections." *Journal of the American Statistical Association* 82 (400): 991–1003.

Spiegelhalter, David, Andrew Thomas, Nicky Best, and Wally Gilks, 1996, "BUGS 0.5: Bayesian Inference Using Gibbs Sampling Manual (Version Ii)." *MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK*, 1–59.

Stevenson, Simon, 2007, "A Comparison of the Forecasting Ability of ARIMA Models." *Journal of Property Investment & Finance*.

UN, 2004, "World Population Prospects: The 2002 Revision. Volume III: Analytical Report." New York.

UN DESA, 2014, "World Urbanization Prospects - Population Division - United Nations." 2014. https://population.un.org/wup/.

UNFPA, 2020, "UNFPA India | United Nations Population Fund." 2020. https://www.unfpa.org/data/transparency-portal/unfpa-india.

Verma, Anurag, Abhinav Singh, Pramendra Pundir, and Gyan Prakash Singh, 2017, "Probabilistic Approach of Population Projection for India and States." In *2017 International Population Conference*. IUSSP.

Whelpton, Pascal K, 1936, "An Empirical Method of Calculating Future Population." *Journal of the American Statistical Association* 31 (195): 457–73.

Zhang, Fengyu, Amy Ong Tsui, and C M Suchindran, 1999, *The Determinants of Contraceptive Discontinuation in Northern India: A Multilevel Analysis of Calendar Data*. MEASURE Evaluation, Carolina Population Center, University of North Carolina ….